19

Towards multimodal interface for interactive robots: challenges and robotic systems description

Burger Brice[1,2,3], Ferrané Isabelle[23] and Lerasle Frederic[13]
[1] CNRS ;LAAS [2] IRIT; UMR CNRS/Université de Toulouse [3] Université de Toulouse; UPS
France

1. Introduction and framework

Making robots to assist people in human-centred environments is a goal to which the robotics community has aspired to for many years (Fong *et al.*, 2003). This field of research is a deep challenge because robots moving out of laboratories have to gain more social skills in order to improve peer-to-peer interaction with a more or less novice user in public, domestic or industrial areas. In contrast to today's specialized service robots these robotic assistants could be well used in such areas and for a variety of tasks like elderly people care, or helping handicapped people as well as assistance in factories or offices. Such prospects require both spatial and transactional intelligence. The former is based on environment perception capabilities. For a robot, this means "being able to understand and navigate in its environment; locating objects and knowing how to manipulate them". The later is based on human perception capabilities. This means "being able to meaningfully communicate with a human user". While the first issue has received much attention in the past, relatively few recent robotic systems are equipped with multimodal user interfaces that permit to control the robot using natural means like human body motion, speech even tactile senses.
RHINO and MINERVA (Thrun *et al.*, 2000) were the first robots to be deployed in a public area, but they do not emphasize the interaction part so much even if they understand speech. Though recent demonstrators embed more advanced human-robot interfaces (frequently based on speech), their capabilities to perceive close human motion remain surprisingly fairly limited (Bennewitz *et al.*, 2005; Maas *et al.*, 2006; Siegwart *et al.*, 2003) or omitted (Breazal *et al.*, 2001; Breazal *et al.*, 2004). To detect persons, ROBOX (Siegwart *et al.*, 2003) relies on two sensors and dedicated algorithms. First, the laser scan data is fed to a motion detector. Secondly, colour images enable the identification of skin colour as well as the subsequent 2D detection and tracking of human faces using simple heuristics.
BIRON (Maas *et al.*, 2006) and ALPHA (Bennewitz *et al.*, 2005) consider also information from heterogeneous sources namely microphone for speaker localization, and vision for frontal face detection... which both lead to intermittent cues. Part of the aforementioned systems (Maas et al., 2006; Siegwart *et al.*, 2003; Thrun *et al.*, 2000) and beyond *i.e.* HERMES

(Bischoff & Graefe, 2004), MAGGIE (Gorostiza et al., 2006), WAKAMARU (Harte & Jarvis, 2007), PEARL (Pineau *et al.,* 2003), etc. focus also on 2D laser data, even radio frequency data *e.g.* ROBOVIE (Kanda *et al.,* 2004) to detect and track humans. To our greater view, using such sensors to perceive human motion seems questionable and vision technologies should be privileged due to their good price-performance ratio and the rich information they encompass.

Perceiving human motion from on-boarded vision is a key-point of human-robot interaction (HRI). On one hand, any interactive robot needs to maintain estimates, or beliefs as probabilistic distributions, about its human user's kinematic (and beyond his/her state) to make effective decisions during interaction. On the other hand, body movements are important in any communication as 65% of the information in a HRI act is non-verbal (Davis, 1971). Visual gestures show human thoughts, replay complements, accent and adjust verbal information. Therefore, vision-based gesture interpretation is valuable in environments where the speech-based communication may be garbled or drowned out. Moreover, the mutual assistance between the robot's speech and vision capabilities enables a user to robustly specify location references in verbal statements. Combined with pointing gestures, such prominent commands open up the possibility of intuitively indicating objects and locations e.g. to make the robot change its direction/position or to mark object. Yet, it can be argued that vision techniques for human perception and natural language processing have mostly been studied rather independently because they constitute research areas in themselves (Pro-danov & Drygajlo, 2003; Skubic *et al.,* 2004; Triesch et al., 2001; Waldherr *et al.,* 2000). Spatial intelligence based on environment perception capabilities has led to numerous and complementary multimodal interfaces in order to label places (Theobalt *et al.,* 2002) or objects (Bischoff & Graefe, 2004) but goes beyond the chapter scope.

Possibly combined with speech for multimodal communication, gesture recognition (GR) has recently received attention in the robotics community. When designing such interface, several requirements must be taken into considerations. First, the visual system must cope with uncontrolled real world environments (background clutter, changes in lighting conditions, presence of several individuals). It must be person independent i.e. many users should be able to operate it, without the necessity for retraining the system. Finally, the system must work at an acceptable speed as on-boarded processing is limited and human users would not accept several seconds for simple gesture recognition tasks. However, hardly any existing interface integrating speech and gesture inputs fulfils all the requirements stated above while symbolic and pointing 3D gestures usually co-occur with speech in a natural interaction scenario. The most advanced multimodal interfaces are probably those presented in (Rogalla *et al.,* 2004; Yoshizaki *et al.,* 2002), even if 2D gestures are considered, and especially (Stiefelhagen *et al.,* 2004) in which the framework trend is similar to ours. In the latter work, a constraint based multimodal system for speech and 3D pointing gestures was developed. Though, monomanual hand gestures are pre-supposed while upper human body extremities are tracked separately, inducing inevitably tracking failures when they overlap. Finally, evaluations are performed on a small data set and independently of any robotic key-scenario. In brief, though remarkable progress has been made, the human motions perceptual capabilities of assistant robots remain fairly limited.

From these considerations, and to best fit these new challenges, a novel framework is proposed here in order to:

- perceive the user's motion namely (i) his/her body placing thanks to the 3D head position which is a commonly stated simplification, (ii) gestures tracking and recognition,
- perform a late-stage fusion with verbal cues to define a multimodal interface,
- integrate and evaluate real-time processes on a mobile robot called JIDO. So, given a targeted scenario, the robustness and the usability of this framework, in the context of a realistic robotic service task is finally investigated.

The tasks we chose are motivated by an "object exchanging" scenario, involving novice human users in the loop: JIDO interprets multimodal orders given by its user in order to move locally, pick a pointed object up, and safely manage to exchange this object with him/her. The primary motivation here is to show that thanks to our interface JIDO is able to interpret multimodal commands like "come to me", "stop", "take this bottle", "bring it to me", "go there", etc. and responds to them by performing the corresponding actions.

In the following, each component integrated on JIDO and involved in the humanrobot interaction process, namely gesture recognition, speech understanding and late stage multimodal fusion are described in section 2. Next, for a global evaluation, live experiments carried out on JIDO in the context of interactive manipulation tasks are described. Finally, this leads us to give some prospects and future work to be done on this topic.

2. Description of JIDO and its multimodal interface

This section gives first some considerations about the JIDO platform and the targeted scenario. Our multimodal interface is embedded on a robot called JIDO which is equipped with a 6-DOF arm, a pan-tilt stereo system at the top of a mast and two laser scanners (Fig. 2). The embedded functionalities are managed thanks to the "LAAS" layered software architecture (Fig. 1) and detailed in (Alami *et al.,* 1998). Such functionalities enable JIDO to:

build maps and navigate in indoor environments. The embedded functionalities are represented in orange boxes.

recognize and manipulate objects thanks to the modules depicted in blue and red boxes. A standard procedure is: the "Hueblob" module extracts the 3D position of the object thanks to blob detection, then "MHP" computes a trajectory which is executed by the "Xarm" module.

perceive humans, namely (i) detection/recognition and view-based tracking from the "ICU" modules, (ii) control of the pan-tilt unit mounted stereo head from the "PTU" module, (iii) 3D gestures tracking and interpretation from the "GEST" modules, (iv) speech utterances interpretation from the "RECO" module, (v) merging of speech and gestures results from the "FUSION" module. The last three modules are briefly described here below.

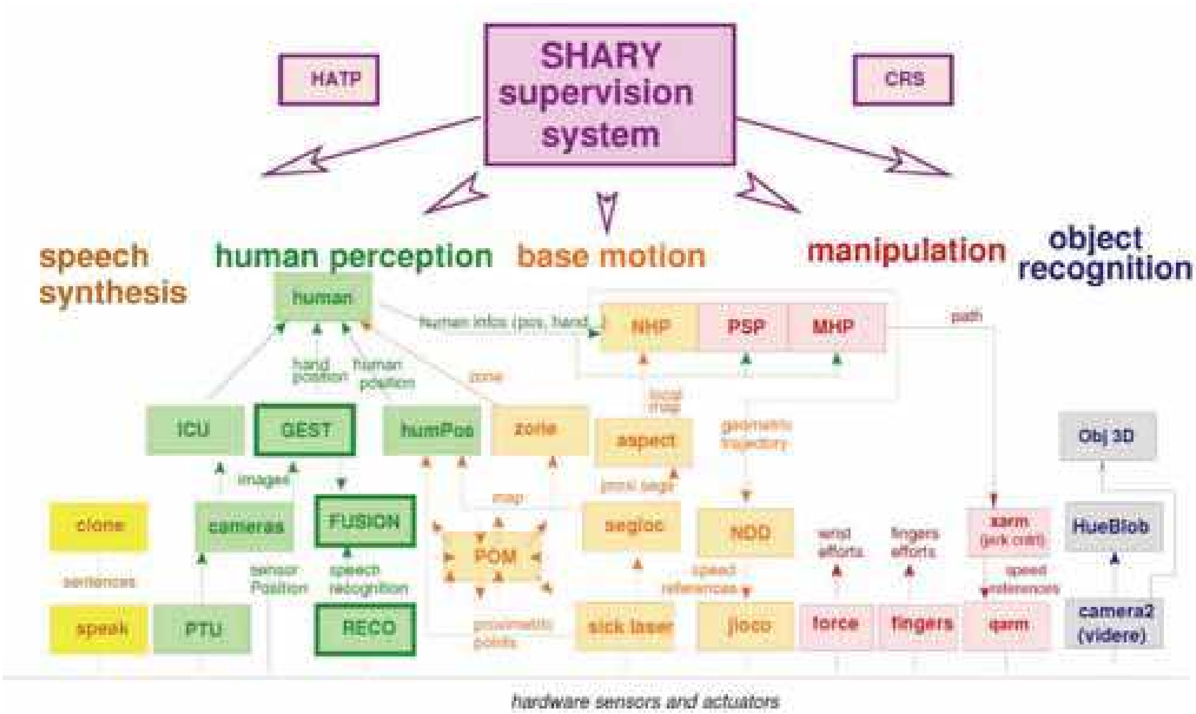


Fig. 1. The JIDO software architecture.

Dynamic gesture recognition is carried out in two phases by the **"GEST" module.** First, the upper extremities of the human body are tracked in 3D in stereoscopic video stream thanks to interactively distributed particle filters devoted to the human's hands and head (Qu *et* al., 2007). Recent investigations concern the second phase, *i.e.* the classification of legitimate gestures. These gestures are here assumed to start and end in the same natural/rest position (the hands lying along the thighs). Given an isolated gesture segment, classification outputs the class the gesture belongs to among a vocabulary composed of: - 7 symbolic gestures defined by their motion templates, namely: "calling out" (with one or two hands), "introducing oneself", "come to me" (with one or two hands), "stop", "go away".
- 5 deictic gestures depending on the coarse pointed direction relatively to the user who performs the gesture *i.e.* "in front of", "bottom left", "bottom right", "top left", "top right". The pointing direction is calculated by the connecting line between the centre of the head and the hand in 3D.
Each gesture is here straightforward modelled by a dedicated HMM. The features used as model's inputs are derived from tracking the 3D positions of both hands relatively to the head to achieve invariance with respect to the person location. For evaluations, we acquired a gesture database consisting of 772 video shots of gestures carried out by 11 different people in front of JIDO. From these experiments, about 70% of the examples are correctly classified. The most prominent error was a failure to recognize "stop" and "introducing oneself" (i.e. "hello") gestures which we can attribute to a poor set of motion template for this gesture. Another observation is that bi-manual gestures are better classified than their mono-manual counterparts as their motion templates are more discriminating.

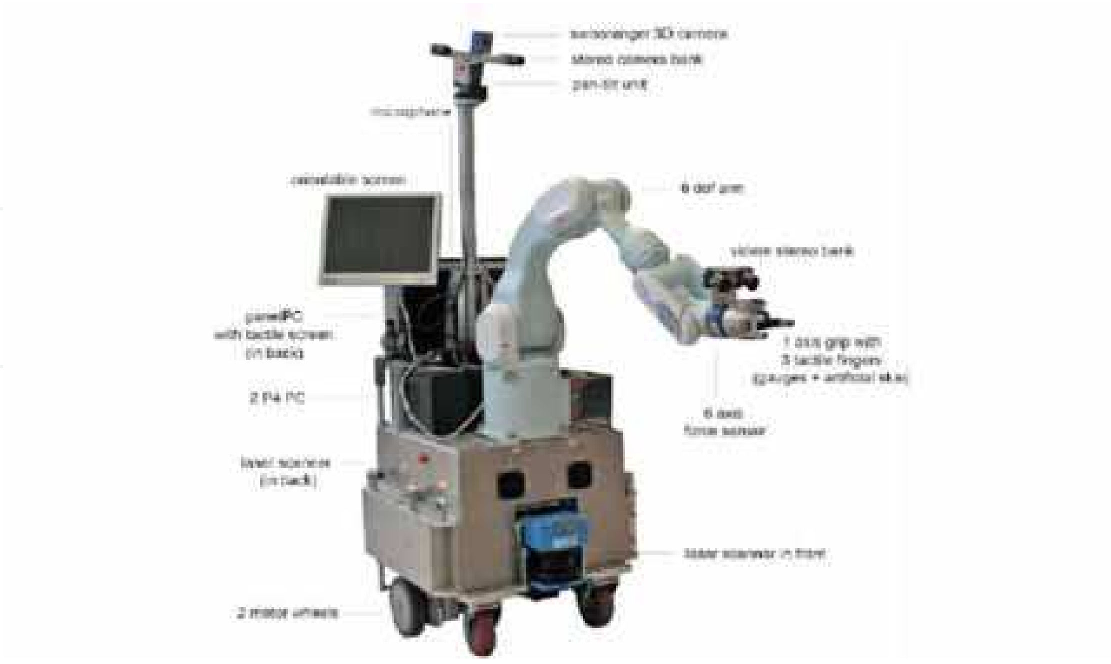


Fig. 2. JIDO and its actuators/sensors.

The "RECO" module performs a classical and speaker-independent speech recognition process, briefly described here. Thirty-nine-parameter vectors are extracted from the audio stream: 12 Mel Frequency Cepstral Coefficients (MFCC), the log energy and their delta and acceleration. The Julius engine is used to decode speech using a set of acoustic and phonetic models for French. This set comprises 37 phonemes and 2 pauses (short/long). It is HMM-based (3-state models with 32 Gaussians per state) and was first trained on about 31 hours of Broadcast News recorded on French radios, in the scope of a completely different transcription task. These models were re-estimated on a thirty-minute audio corpus, directly recorded on the robotic platform in the same conditions as our other experiments. The lexicon (246 words/428 pronunciations drawn up from the French lexical database BDLEX (Perennou & De Calmes, 2000) and the language model were specifically designed for our multimodal experiments. Deictic and anaphoric sentences as well as some language flexibility were taken into account. Context free grammars were designed for different types of user requests like starting interaction with the robot, asking for guidance, global robot or robot's arm movements, or object exchanges. This represents an overall set of 2334 different well-formed sentences, although the experiments described in the robotic scenario only involved a little subset of them, focusing on local movement and deictic requests for object exchange. Nevertheless, the evaluations carried out in our experimental conditions were made on 1200 various sentences, covering all request types and uttered by 16 different speakers (including 7 non native). The Word Error Rate (WER) is around 7%, while the Sentence Error Rate (SER) is equal to 19%. As word recognition errors have an immediate impact on sentence recognition and a less immediate one on sentence interpretation, when errors occur on words that are not significant in the interaction context, the sentence Interpretation Error Rate (IER) is 6% lower (13%) than the SER.
The "FUSION" module merges gesture recognition results and speech interpretation thanks to late-stage and hierarchical fusion ones. The speech is used as the main channel and

actions needing a gesture disambiguation are identified by the "RECO" module. Following a rule-based approach, the command generated by "RECO" is completed. Thus, for human-dependent commands *e.g.* "viens ici" ("go there"), the human position and the pointed direction are characterized thanks to the 3D visual tracker. Late-stage fusion consists of fusing the confidence scores for each N-Best hypothesis produced by the speech and vision modalities according to (Philipp *et al.,* 2008). The associated performances are reported thanks to the targeted robotic scenario detailed here below.

3. Targeted scenario and robotics experiments

These "human perception" modules encapsulated in the multimodal interface have been undertaken within the following key-scenario (Table 1). Since we have to deal with robot's misunderstandings, we refer to the human-human communication and the way to cope with understanding failure. In front of such situations, a person generally resumes his/her latest request in order to be understood. In our scenario, although no real dialogue management has been implemented yet, we wanted to give the robot the possibility to ask the user to repeat his/her request each time one of the planed step fails without irreversible consequences. By saying "I did not understand, please try again." *(via* the speech synthesis module named "speak"), the robot resume its latest step at its beginning. The multimodal interface runs completely on-board the robot. From this key-scenario, several experiments were conducted by several users in our institute environment. They asked JIDO to follow their instructions given by means of multimodal requests: by first asking JIDO to come close to a given table, take over the pointed object and give it to him/her. Figure 3 illustrates the scenario execution. For each step, the main picture depicts the current H/R situation, while the sub-figure shows the tracking results of the GEST module. In this trial, the multimodal interface succeeds to interpret multimodal commands and to safely manage objects exchanges with the user.

#	The human user command	The JIDO action	Demonstrated modules	Comments
1.	"Hello, I am here" accompanied with a symbolic gesture	Local motion towards the user	RECO, GEST, FUSION, MHP, PSP	JIDO moves and stops in front of the user
2.	"Hi JIDO it's Brice"		RECO, ICU	The user must be identified to be granted to interact with JIDO
3.	"Come to me" with a pointing gesture	Location motion towards the pointed location on the floor	GEST, RECO, FUSION, MHP	The command execution requires the 3D location of the user
4.	"Stop" with a symbolic gesture"	Stop of the robot	GEST, RECO, FUSION	This command is performed while the robot is moving
5.	"Take this object" with a pointing gesture	Grasping of the object	GEST, RECO, FUSION, Hue-Blob, MHP, Xarm	Jido searches for an object the user points to, then picks it up
6.	"Go to my left side" with a pointing gesture	Local motion according to the user location	GEST, RECO, FUSION, MHP, PSP	The command execution requires the 3D location of the user
7.	"Give the object to me"	Object handling	GEST, RECO, FUSION, MHP, Xarm	The command execution requires no gesture recognition but only the hand tracking
8.	"Go away" with a symbolic gesture	Local motion to go away from the user	GEST, RECO, FUSION, MHP	

Table 1. Excerpt of an interaction scenario between a human user and JIDO.

Fig. 3. From top-left to bottom-right, snapshots of a scenario involving speech and gesture recognition and data fusion: current H/R situation -main frame-, "GEST" module results -bottom right then bottom left-, other modules ("Hue Blob", "ICU") results -top-.

Given this scenario, quantitative performance evaluations were also conducted. They refer to both (i) the robot capability to execute the scenario, (ii) and potential user acceptance of the ongoing interaction scenario. The less failures of the multimodal interface will occur, the more comfortable the interaction act will be for the user. The associated statistics are summarized in Table 2 which synthesizes the data collected during 14 scenario executions.
Let us comment these results. In 14 trials of the full scenario execution, we observed only 1 fateful failure (noted fatal) which was due to a localisation failure and none attributable to our multimodal interface. Besides, we considered that a run of this scenario involving more than 3 failures is potentially unacceptable by the user, who can be easily bored by being constantly asked to re-perform his/her request. These situations were encountered when touching the limits of our system like for example when the precision of pointing gestures decreases with the angle between the head-hand line and the table. In the same manner, short utterances are still difficult to recognize especially when the environment is polluted with short sudden noises.

#	"RECO"	"GEST"	"FUSION"	Others	Comments
1.	0	1	0	0	
2.	0	0	0	1 ICU	Face recognition
3.	1	3	1	0	The distance to the robot makes this gesture hard to track
4.	3	2	2	0	Computing time sometimes too long when the robot is moving
5.	0	0	0	2 HueBlob	The bottle is not always seen
6.	0	0	0	0	The left is not really on the left...
7.	0	0	0	2 MHP (1 fatal)	Hand too far, localisation failure
8.	2	4	1	0	

Table 2**.** Modules' failure rates during scenario trials.

Apart from these limitations, the multimodal interface is shown to be robust enough to allow continuous operation for the long-term experimentations that are intended to be performed.

4. Conclusion

This article described a multimodal interface for a more natural interaction between humans and a mobile robot. A first contribution concerns gesture and speech probabilistic fusion at the semantic level. We use an open source speech recognition engine (Julius) for speaker independent recognition of continuous speech. Speech interpretation is done on the basis of the N-best speech recognition results and a confidence score is associated with each hypothesis. By this way, we strengthen the reliability of our speech recognition and interpretation processes. Results on pre-recorded data illustrated the high level of robustness and usability of our interface. Clearly, it is worthwhile to augment the gesture recognizer by a speech-based interface as the robustness reaches by cue proper fusion is much higher than for single cues. The second contribution concerns robotic experiments which illustrated a high level of robustness and usability of our interface by multiple users. While this is only a key-scenario designed to test our interface, we think that the latter opens in increasing number of interaction possibilities. To our knowledge, quite few mature robotic systems enjoy such advanced embedded multimodal interaction capabilities.

Several directions are currently studied regarding this multimodal interface. First, our tracking modality will be made much more active. Zooming will be used to actively adapt the focal length with respect to the H/R distance and the current robot status. A second envisaged extension is, in the vein of (Richarz *et al.,* 20006; Stiefelhagen *et al.,* 2004), to incorporate the head orientation as additional features in the gesture characterization as our robotic experiments strongly confirmed by evidence that a person tends to look at the pointing target when performing such gestures. The gesture recognition performances and the precision of the pointing direction should be increased significantly. Further investigations will aim to augment the gesture vocabulary and refine the fusion process, between speech and gesture. The major computational bottle-neck will become the gesture recognition process. An alternative, pushed forward by (Pavlovic *et al.*, 1999), will be to privilege dynamic Bayesian networks instead of HMMs which implementation requires linear increasing complexity in terms of the gesture number.

5. Acknowledgements

The work described in this chapter was partially conducted within the EU Project CommRob "advanced Robot behaviour and high-level multimodal communication" (URL www.commrob.eu) under contract FP6-IST-045441.

6. References

Alami, R.; Chatila, R.; Fleury, S.; Ingrand, F. (1998). An architecture for autonomy. *Int. Journal of Robotic Research (IJRR'98),* 17(4):315—337.

Benewitz, M.; Faber, F.; Joho, D., Schreiber, M.; Behnke, S. (2005). Towards a humanoid museum guide robot that interacts with multiple persons. *Int. Conf. on Humanoid Robots (HUMANOID'05),* pages 418—423, Tsukuba, Japan.

Bischoff, R.; Graefe, V. (2004). HERMES - a versatile personal robotic assistant.*IEEE,* 92:1759—1779.

Breazal, C.; Brooks, A.; Chilongo, D.; Gray, J.; Hoffman, A.; Lee, C.; Lieberman, J. (2004). Working collaboratively with humanoid robots. *ACM Computers in Entertainment,* July.

Breazal, C.; Edsinger, A.; Fitzpatrick, P.; Scassellati, B. (2001). Active vision for sociable robots. Trans. *On Systems, Man, and Cybernetics,* 31:443—453.

Davis, F (1971). Inside Intuition What we know about non-verbal communication. *Mc Graw-Hill book Co.*

Fong, T; Nourbakhsh, I.; Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems (RAS'03),* 42: 143—166.

Gorostiza, J.; Barber, R.; Khamis, A.; Malfaz, M. (2006). Multimodal human-robot Framework for a personal robot. *Symp. on Robot and Human Interactive Communication (RO-MAN'06),* pages 39—44, Hatfield, UK.

Harte, E.; Jarvis, R. (2007). Multimodal human-robot interaction in an assistive technology context. *Australian Conf. on Robotics and Automation,* Brisbane, Australia.

Isard, M.; Blake, A. (1998). I-CONDENSATION: unifying low-level and high-level tracking in a stochastic framework. *European Conf. on Computer Vision (ECCV'98),* pages 893—908, Freibourg, Germany.

Kanda, T.; Ishiguro, H.; Imai, M.; Ono, T. (2004). Development and evaluation ofInteractive humanoide robots. *IEEE,* 92(11): 1839—1850.

Maas, J.F.; Spexard, T.; Fritsch, J.; Wrede, B.; Sagerer, G. (2006). BIRON, what's the topic ? A multimodal topic tracker for improved human-robot interaction. *Symp. on Robot and Human Interactive Communication (RO- MAN'06),* Hatfield, UK.

Pavlovic, V.; Rehg, J.M.; Cham, T.J. (1999). A dynamic Bayesian network approach to tracking using learned switching dynamic models. *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'99),* Fort Collins, USA.

Pérennou, G.; De Calmes, M. (2000). MHATLex: Lexical resources for modeling the french pronunciation. *Int. Conf. on Language Resources andEvaluations,* pages 257—264, Athens, Greece.

Philipp, W.L.; Holzapfel, H.; Waibel, A. (2008). Confidence based multimodal fusion for person identification. *In ACM Int. Conf. On Multimedia*, pages 885-888, New York, USA.

Pineau, J.; Montemerlo, M.; Pollack, M.; Roy, N.; Thrun, S. (2003). Towards robotic assistants in nursing homes: challenges and results. *Robotics and Autonomous Systems (RAS'03),* 42:271—281.

Prodanov, P.; Drygajlo, A. (2003). Multimodal interaction management for tour-guide robots using Bayesian networks. *Int. Conf. on Intelligent Robots and Systems (IROS'03),* pages 3447—3452, Las Vegas, USA.

Qu, W.; Schonfeld, D.; Mohamed, M. (2007). Distribution Bayesian multiple-targettracking in crowded environments using collaborative cameras. *EURASIP Journal on Advances in Signal Processing.*

Rabiner, L. (1989). A tutorial on hidden markov models and selected applicationsin speech recognition. *IEEE,* 77(2): 257—286.

Richarz, J.; Martin, C.; Scheidig, A., Gross, H.M. (2006). There you go ! Estimatingpointing gestures in monocular images from mobile robot instruction.*Symp. on Robot and Human Interactive Communication (RO-MAN'06),*pages 546—551, Hartfield, UK.

Rogalla, O.; Ehrenmann, M.; Zollner, R.; Becher, R.; Dillman, R. (2004). Usinggesture and speech control for commanding a robot. *Book titledAdvances in human-robot interaction,* volume 14, Springer Verlag.

Siegwart, R.; Arras, O.; Bouabdallah, S.; Burnier, D.; Froidevaux, G. ; Greppin, X. ;Jensen, B. ; Lorotte, A. ; Mayor, L. ; Meisser, M. ; Philippsen, R. ; Piguet,R. ; Ramel, G. ; Terrien, G., Tomatis, N. (2003). RoboX at expo 0.2: alarge scale installation of personal robots. *Robotics and Autonomous Systems (RAS'03),* 42:203—222.

Skubic, M.; Perzanowski, D.; Blisard, S.; Schutz, A.; Adams, W. (2004). Spatial language for human-robot dialogs. Journal of Systems, *Man andCybernetics,* 2(34):154—167.

Stiefelhagen, R.; Fugen, C.; Gieselman, P.; Holzapfel, H.; Nickel, K., Waibel, A. (2004). Natural human-robot interaction using speech, head pose and gestures. *Int. Conf. on Intelligent Robots and Systems (IROS'04),* Sendal, Japan.

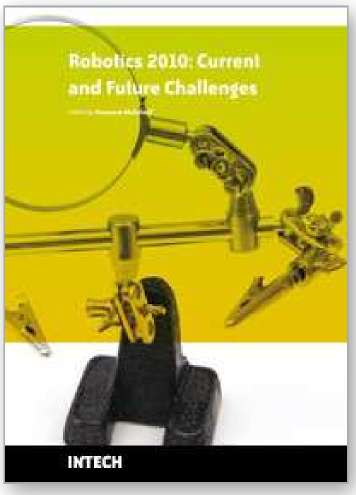
Theobalt, C.; Bos, J.; Chapman, T.; Espinosa, A. (2002). Talking to godot: dialogue with a mobile robot. *Int. Conf. on Intelligenr Robots and Systems (IROS'02),* Lausanne, Switzerland.

Thrun, S.; Beetz, M.; Bennewitz, M.; Burgard, W.; Cremers, A.B.; Dellaert, F.; Fox, D.; Halnel, D.; Rosenberg, C.; Roy, N.; Schulte, J. Schulz, D. (2000). Probabilistic algorithms and the interactive museum tour-guide robot MINERVA. *Int. Journal of Robotics Research (IJRR'00).*

Triesch, J.; Von der Malsburg, C. (2001). A system for person-independent hand posture recognition against complex background. *Trans. On Pattern Analysis Machine Intelligence (PAMI'01),* 23(12):1449-1453.

Waldherr, S.; Thrun, S.; Romero, R. (2000). A gesture-based interface for humanrobot interaction. *Autonomous Robots (AR'00),* 9(2): 151—173.

Yoshizaki, M.; Kuno, Y.; Nakamura, A. (2002). Mutual assistance between speech and vision for human-robot interface. *Int. Conf. on Intelligent Robots and Systems (IROS'02)*, pages 1308—1313, Lausanne, Switzerland.



Robotics 2010 Current and Future Challenges
Edited by Houssem Abdellatif

ISBN 978-953-7619-78-7
Hard cover, 494 pages
Publisher InTech
Published online 01, February, 2010
Published in print edition February, 2010

Without a doubt, robotics has made an incredible progress over the last decades. The vision of developing, designing and creating technical systems that help humans to achieve hard and complex tasks, has intelligently led to an incredible variety of solutions. There are barely technical fields that could exhibit more interdisciplinary interconnections like robotics. This fact is generated by highly complex challenges imposed by robotic systems, especially the requirement on intelligent and autonomous operation. This book tries to give an insight into the evolutionary process that takes place in robotics. It provides articles covering a wide range of this exciting area. The progress of technical challenges and concepts may illuminate the relationship between developments that seem to be completely different at first sight. The robotics remains an exciting scientific and engineering field. The community looks optimistically ahead and also looks forward for the future challenges and new development.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Burger Brice, Ferrane Isabelle and Lerasle Frederic (2010). Towards Multimodal Interface for Interactive Robots: Challenges and Robotic Systems Description, Robotics 2010 Current and Future Challenges, Houssem Abdellatif (Ed.), ISBN: 978-953-7619-78-7, InTech, Available from: http://www.intechopen.com/books/robotics-2010-current-and-future-challenges/towards-multimodal-interface-for-interactive-robots-challenges-and-robotic-systems-description

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821